Exhibit B

ELSEVIER

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Biochimica et Biophysica Acta 1382 (1998) 191–206

Review

The structural aspects of limited proteolysis of native proteins

Simon J. Hubbard *

Department of Biomolecular Sciences, University of Manchester, Institute of Science and Technology, P.O. Box 88, Manchester M60 1QD, UK

Received 11 July 1997; revised 3 October 1997; accepted 3 October 1997

Contents

* Corresponding author. Fax: (44) (0)161 236 0409; E-mail: sjh@sjh.bi.umist.ac.uk

PII S0167-4838(97)00175-1

1. Introduction

Limited proteolysis is responsible for activating a wide range of proteins from immature forms and is h nce implicated in a number of biologically important systems [1–3]. Proteases have also been used widely throughout the field of biochemistry in many areas of study including sequencing, enzyme (in)activation and complete degradation [4,5]. Often these hydrolytic reactions take place under denaturing conditions, so that every peptide bond is cut where the local amino acid sequence satisfies the specificity requirements of the protease in question. This application is useful for determination of amino acid sequence, such as the generation of a complete limit digest for mass spectrometric analysis. The protease trypsin is now routinely used to produce such a limit "map" for protein identification. However, the denaturation step precludes any information being gained on the three-dimensional structure of the protein. In order to retrieve structural information through proteolytic digestion it is necessary to limit the reaction in some way, leading to the notion of "limited" proteolysis. A classic example of a limited proteolysis is the cleavage of ribonuclease A by subtilisin at a bond between residues 20 and 21. The product, ribonuclease S, is made up of the non-covalently associated S-peptide and S-protein and is fully active [6].

Limitation of the proteolytic reaction can be achieved in a number of ways, such as by alteration of the reaction conditions (temperature, pH, ionic strength), or more typically, by ensuring the substrate protein is in a native (or near native) state. If this last condition is satisfied, limited proteolysis is a powerful tool for probing the higher order structure of proteins; providing information on the location of particular peptide bonds with respect to the overall fold of the protein. The premise underpinning such studies is the sequence/structure paradigm of limited proteolysis: that higher order structure and not primary sequence is the main determinant of the site of initial hydrolysis. For example, trypsin cleaves peptide bonds C-terminal to basic amino acids, lysine and arginine. Assuming that 1 in 10 of all amino acids in a protein sequence are lysine or arginine, 10% of all peptide bonds are in principle accessible to proteolytic attack and subsequent cleavage by this enzyme. However, under non-denaturing conditions this is rarely the case and typically only a few, one or sometimes even none, of the putative bonds are cut in the time scale of a typical experiment. The reasons for this are obviously structural. Peptide units buried in the protein core and in regular secondary structural elements are less accessible to the enzyme and are therefore not cleaved as quickly. Hence, a priori, it would be expected that the observed cleavages occur at broadly "surface" sites, that are exposed to the surrounding solvent, such as loops. This assumption, borne out by experimental observation, has been widely exploited to infer the accessibility of sites in proteins of unknown structure. The same premise applies to the identification of domain linking segments which are similarly expected to be surface exposed and readily cleaved by proteinases. Knowledge of the surface amino acid residues is clearly of great benefit for understanding protein function as it can inform on likely epitopes and receptor-interaction sites. This general approach has also now been extended to the study of near-native partly unfolded states, to gain information on protein folding and its pathways [7].

Although this technique provides a simple, cheap and effective method for detecting exposed surface sites, there remain caveats to the model. Specifically, is it really only exposure that dictates the site of cleavage for a protease of broad specificity? A number of research groups, including our own, have studied limited proteolysis as a molecular recognition system to provide a more precise understanding of the structural determinants of this process. Factors other than surface exposure must also act as key determinants of vulnerability of proteolysis. Clearly, a more detailed understanding of the molecular events involved in the recognition process is required in order to use limited proteolysis as a thoroughly reliable probe of protein structural analysis.

This review will focus on limited proteolysis as a tool for the investigation of protein structure, using recent exemplary studies to highlight its utility, as well as pointing out the potential pitfalls. The underlying structural determinants will be discussed by reference to a number of recent studies on proteins of known structure, as well as a statistical survey of the structural properties of limited proteolytic sites (nicksites) and attempts to predict them a priori from structure and sequence. Finally, future directions for

this research area will be address d, including the use of limited proteolysis in the study of protein folding and potential biot ch applications.

2. Th sequence-structure paradigm of limited proteolysis

The basic paradigm of limited proteolysis has already been described: namely, the masking of primary specificity by tertiary structure. It is worth briefly describing the origins of primary specificity. The earliest systems to be structurally characterised were the serine proteinases chymotrypsin and trypsin [8,9]. Their X-ray crystal structure determinations allowed the known amino acid sequence specificities to be rationalised in terms of subsite binding pockets located at the enzyme active site cleft [10]. The subsites are classified from the notation of Schechter and Berger [11] by S_n which corresponds to the side chain of the substrate polypeptide P_n which binds into it. The side chains are numbered from P_1 on the amino-terminal side of the scissile peptide and P'_1 on the carboxy-terminus. Hence the primary specificity of the enzyme is usually defined in terms of the P_1 side chain preference for the S_1 subsite. In trypsin, a deep S_1 site is formed that possesses an acidic aspartic acid (Asp189) at the pocket base. This gives trypsin a strong preference for the basic amino acids lysine and arginine at P_1. In fact, under normal conditions trypsin will cleave at no other bond. In chymotrypsin, the equivalent position to this acidic side chain is substituted by a neutral residue, serine, and the hydrophobic properties of the pocket become dominant. This leads to a preference for hydrophobic (and preferably) aromatic side chains. In elastase, the pocket is partly filled by the mutation of two glycine residues in trypsin and chymotrypsin at Gly214 and Gly226 to valine and threonine, respectively, leading to a hindrance of all but the small amino acids such as alanine and serine at the P_1 position. Recently, Laskowski and co-workers have attempted to quantify the effect of all of the 20 commonly occurring amino acids at P_1 for a core of six well known serine proteases [12]. This was achieved by recombinant techniques on the ovomucoid inhibitor from turkey. Binding studies yielded kinetic data which were converted for free energies to give a quantitative scale for all the amino acids at the primary recognition position for these six enzymes. In general, the scales were unsurprising and might have been predicted, at least qualitatively. The hydrophobic interaction was observed to be the dominant force, moderated by size and shape effects and polarity. However, it was also clear that minor alterations in the pocket geometry could have a large effect on the overall preference of even very closely related enzymes such as human leucocyte and pancreatic elastase [12,13].

In addition to the primary specificity, some enzymes exhibit a secondary specificity at sites removed from the scissile bond. Subtilisin shows a strong preference for hydrophobic side chains at the P_4 position [14]. In subtilisin BPN′ the amino acids Tyr104, Ile107 and Leu126 create a hydrophobic pocket that demonstrates a preference for the following amino acid side chains at this position: Phe > Leu, Ile, Val > Ala [15]. Other secondary subsites have been discovered in other proteinases although their effects on binding and catalysis are usually small.

Despite these primary and secondary subsite preferences, subtilisin will readily cleave after almost any amino acid. In some cases, the subsite specificity of the enzyme therefore becomes almost irrelevant. In the context of limited proteolysis, the proteinases may be classed into two categories: those with very narrow specificity requirements such as trypsin, V8-proteinase and endoproteinase Arg-C, and those with relatively broad specificity requirements such as subtilisin, thermolysin and proteinase K. Although sequence preferences must play a role in dictating the site of ultimate hydrolysis in the folded protein, the influence of tertiary structure is expected to be more pronounced for the broader specificity proteinase.

3. Experimental considerations

Reaction conditions for limited proteolysis are typically chosen to ensure that complete degradation of the substrate protein does not take place. In order to limit the digestion, a number of methods are typically employed. Often, the enzyme:substrate ratio is restricted to somewhere between 1:50 and 1:1000 so that proteolysis is incomplete and intermediates may be observed accumulating over time. Additional techniques include performing the digest at low tempera-

ture and non-optimal pH for the enzyme. However, some enzyme–substrate systems require no retardation as the protein substrate is strongly resistant to proteolysis. In these instances, it may be necessary to use higher enzyme concentrations, elevated temperatures, extended digest times or even low concentrations of denaturing agents. In the latter case, it is important from the point of view of structural inference, that the tertiary fold of the protein has not been greatly distorted. No significant loss of function or activity should be observed and the protein should possess the same structural properties. Generally, th re are no hard and fast rules governing enzyme:substrate ratios and optimal experimental conditions to study limited proteolysis must be found out by preliminary experimentation.

Most proteolytic reactions are monitored via SDS-PAGE, a simple and cheap technique available to most laboratories. However, gel electrophoresis of proteinase digests will rarely yield the precise site of limited proteolysis unless the proteinase is particularly narrow in its primary specificity and the amino sequence fortuitously disposed. To determine the exact site(s) of proteolysis, further studies are required such as Edman degradation sequencing chemistry. Mass spectrometric methods are also being used with increasing frequency due to the high mass accuracy they are able to yield, particularly with low sample quantities [16,17].

4. Structural determinants of limited proteolysis

Early studies on limited proteolysis considered the reaction in the context of the compact, globular structure of folded proteins and Linderström–Lang proposed two pathways for degradation dependant on the stability of the protein [18]. The initial nicking may destroy the stability of the protein, which subsequently unfolds, exposing all peptides to attack and the protein is degraded to completion. In the second case, the protein retains its overall structure and remains folded, preventing general degradation from taking place. In the light of subsequent high resolution X-ray crystal structure determinations of known substrate proteins, Neurath proposed the original hypothesis that limited proteolysis occurs at "hinges and fringes" such as exposed surface loops and domain linking segments [4]. These theories were further expanded by other workers who attempted to quantify the contributions of exposure and flexibility to limited proteolysis. One study considered the known autolytic and subtilisin cleavage sites in the protein thermolysin, itself a proteinase [19,20]. A plot of the residue-averaged atomic temperature factors of thermolysin obtained from a crystallographic determination demonstrated that the limited proteolytic sites exhibited a remarkable correlation with the peaks in the profile. Temperature factors are in part derived from the thermal motions that occur in the crystalline protein molecule and represent a useful measure of the segmental mobility of the substrate amino acid chain. Hence, the limited proteolytic sites in thermolysin occurred in flexible regions of the molecule. An example of this correlation is shown in Fig. 1 for the protein staphylococcal nuclease which is cleaved by trypsin at Lys48–Lys49 and Lys49–Gly50 in non-denaturing conditions [21,22]. The profile shows that these two trypsin nicksites are situated at th largest peaks and are therefore the most flexible putative tryptic sites (as characterised by temperature factors).

An opposing hypothesis to these flexibility theories was obtained by using a large spherical probe as a model of a proteinase and calculating the accessible surface when rolled around the surface of three protein structures [23]. This data suggested that surface exposure and not flexibility was the prime determinant of limited proteolysis. However, subsequent studies of an increased number of proteins have not

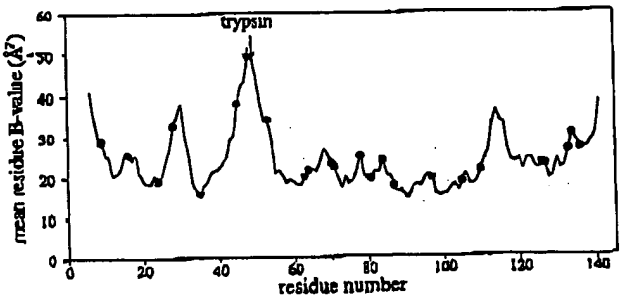


Fig. 1. Temperature factor profile for Staphylococcal nuclease. Mean residue temperature factors (averaged over all backbone atoms) are plotted al ng the amino acid sequence of the protein. All putative trypsin nicksites are indicated by filled circles on the profile and the actual nicksites cut by the enzyme are indicated by arrows.

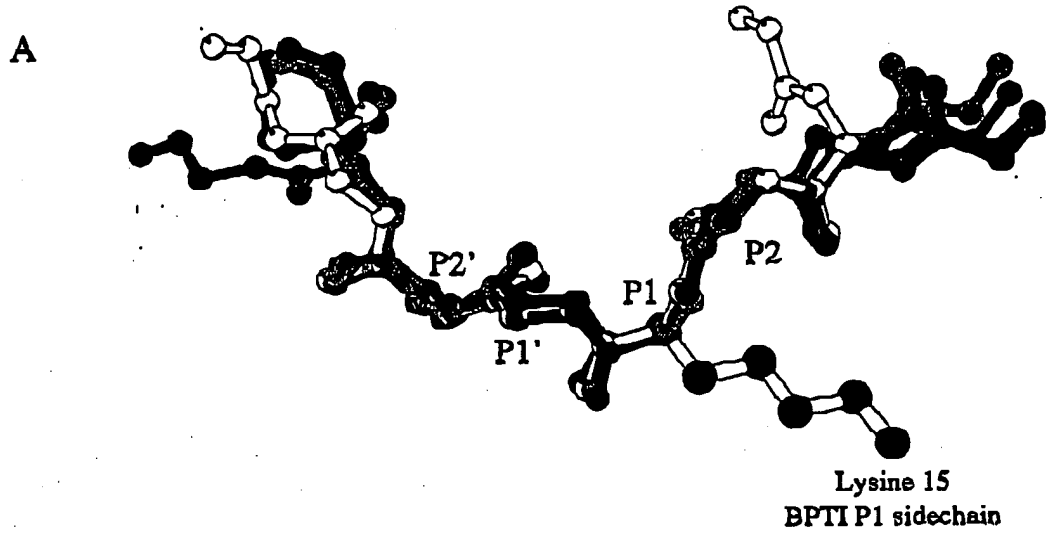
reproduced these findings. Fontana, when reviewing the thermolysin data [24,25], expanded the considerations to other systems and concluded again that segmental mobility of the protein chain is the overriding factor in determining susceptibility.

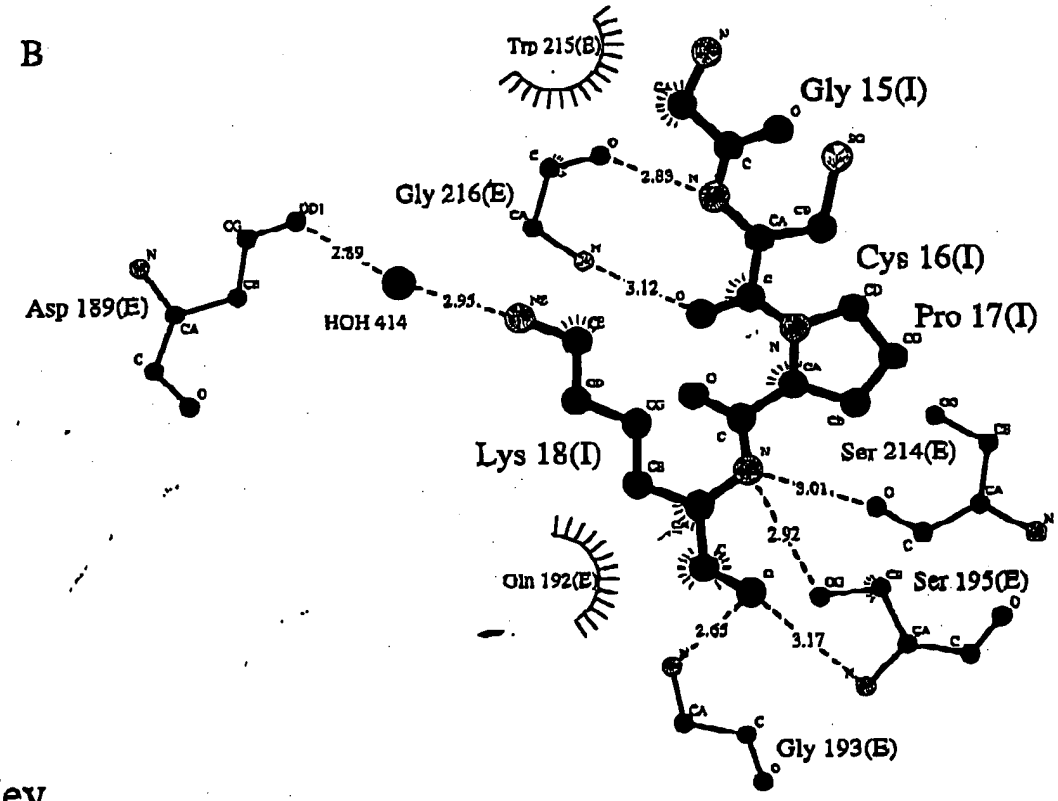
Other studies have considered the precise conformation of the substrate regions of nicked proteins more carefully [26]. At some stage prior to cleavage, the substrate peptide must closely match the structure of the "idealised substrate" conformation exhibited by the canonical binding-loop conformation observed in the families of small protein inhibitors. When compared one to another, the protein inhibitor binding loops were seen to possess practically identical backbone conformations spanning the P_2–P'_2 region [26,27]. A superposition of some representative serine proteinase inhibitor binding loops is shown in Fig. 2(a) depicting the high degree of structural backbone conservation. In Fig. 2(b), the conserved interactions made by these inhibitor structures to the enzyme are schematised using the pancreatic secretory trypsin inhibitor (PSTI)/trypsin complex as representative. The peptide carbonyl carbon is approached by the attacking catalytic serine of the proteinase, whilst the P_1 side chain binds into the S_1 pocket. The carbonyl oxygen is bound by the amide groups of enzyme residues Gly193 and Ser195 which form the so-called oxyanion binding pocket. These interactions, or their equivalents, would be expected to be made prior to cleavage for all serine protease substrates.

Tryptic limited proteolytic sites possess totally different structures to the conserved inhibitor template, and were also different to each other [26]. Despite this, the nicksites were also generally well correlated with accessibility, protrusion and flexibility (as characterised by X-ray crystallographic temperature factors). However, from inspection and simple modelling experiments, it was clear that a structural rearrangement would be required, primarily a local unfolding step, in order for proteolysis to take place. This conclusion was confirmed by a series of loop modelling experiments designed to test this theory and attempt to quantify the degree of local unfolding required [28]. Plausible substrate models for known proteolytically cleaved loops were only achieved when upwards of 10 amino acids were allowed to unfold locally. Similarly, the modelling protocol demonstrated that β-strand was unsuited to hosting a limited proteolytic site although they were plausible within α-helices. This is illustrated in Fig. 3, where a loop containing the required template conformation from P_2–P'_2 may be introduced into a helical segment but not in a beta (extended region). The rationale for this is that too many inter-strand hydrogen bonds would need to be broken and the ends brought closer together to deform an extended chain segment such as found in β-sheets. The deformation of extended segments such as β-strands is necessary to allow the nicksite region to fit into the enzyme active site without introducing a large number of unfavourable contacts with the lip of the cleft and the rest of the enzyme. Although there are less geometrical restrictions within α-helices, similar energetic constraints are believed to disfavour the location of nicksites in the centre of helices [7]. Indeed, this disposition of nicksites to occur in regions of non-secondary structure is widely observed.

In addition to the modelling of known nicksites, all the putative tryptic sites in elastase were passed through the same modelling procedure. This produced the interesting result that the most easily modelled site was the true nicksite at Arg125–Ala126 [29]. Successful models could only be produced for a handful of the possible sites. A re-modelled loop was deemed acceptable if it contained the requisite conformation from P_2–P'_2, did not make unfavourable interactions with the rest of the protein bulk and could be "docked" into the active site of the enzyme without introducing a large number of unfavourable intermolecular contacts. Of the 4 or 5 sites where this was possible, the sites were also considered by the number of hydrogen bonds each loop made to the rest of the protein and the amount of surface area they buried. The true nicksite was found to make the fewest hydrogen bonds with the rest of the protein and buried relatively little surface. Thus, the most susceptible site was characterised by making the fewest interactions with the rest of the protein and is thus the most readily locally unfolded.

Although conformational parameters such as accessibility, segmental mobility and protrusion correlate quite well with limited proteolytic sites, these parameters are themselves highly correlated [30]. As pointed out by Fontana [7], this in itself is not enough to rationalise the limited proteolysis phenomenon, as it is not clear which of these factors are important





Key

PSTI bond

Trypsinogen bond

Hydrogen bond and its length

His 53 Non-ligand residues involved in hydrophobic contact(s)

Corresponding atoms involved in hydrophobic contact(s)

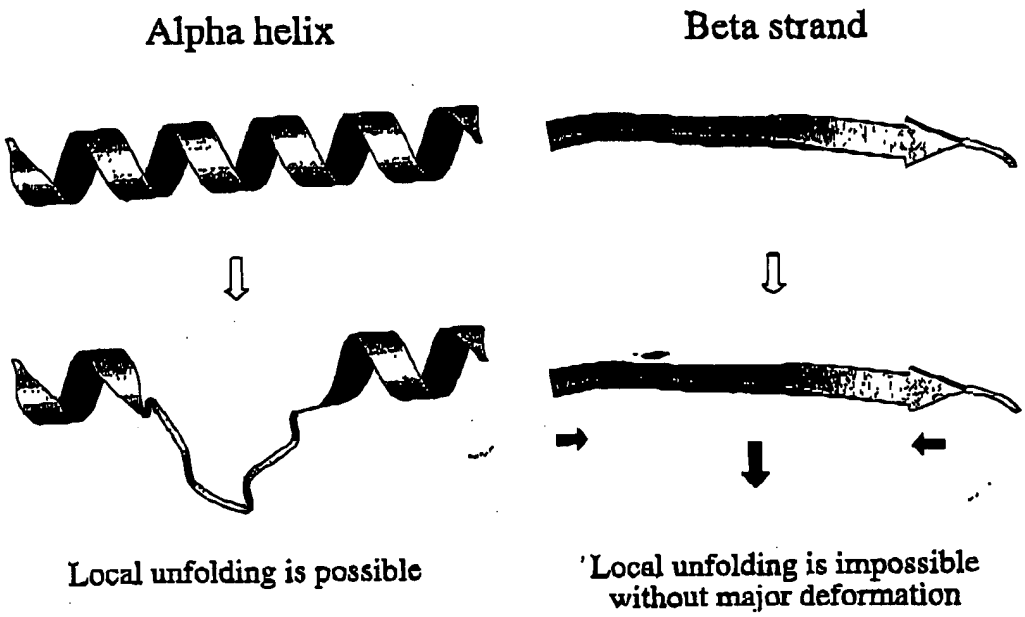


Fig. 3. Secondary structural models for local unfolding. The two regular secondary structure states, α-helix and β-strand, are depicted as MOLSCRIPT [77] cartoons. Loop closure modelling experiments were able to show that local unfolding could occur with regular helix although it was not geometrically possible within extended β-structure without gross deformation of the protein structure.

and which merely correlated. Indeed, even when a protein structure is known to high resolution atomic detail and its dynamic properties characterised, it is not always possible to explain proteolysis results. What remains unchallenged is the need for local unfolding to take place, and it seems most likely that the regions of protein structure most able to do this will be those characterised by the conformational parameters discussed here.

5. Limited proteolysis as a structural probe

The incomplete understanding of the factors governing limited proteolysis notwithstanding, it is still widely used as a probe for protein structure [31]. This is based on the early hypotheses that limited proteolysis occurs exclusively at "hinges and fringes" [4]. In the case of native globular protein structures the reaction is expected to occur primarily at flexible surface loops and in the case of multi-domain proteins at similarly mobile linker regions between the domains. The hydrolysis typically yields a nicked species that retains its overall fold under non-denaturing conditions for most single domain proteins and potentially separates the individual domains for multi-domain proteins. Hence limited proteolysis may be used to monitor protein surface regions, ligand-induced conformational changes, domain boundaries and protein unfolding/refolding. Some examples of

Fig. 2. Stereochemistry at the recognition regions during limited proteolysis. In A, a MOLSCRIPT [77] carton of the structural superposition of the binding loops from P_4–P'_4 of five serine protease inhibitors is shown. Only the backbone atoms are shown for clarity apart from the P_1 side chain of lysine 15 from BPTI. The inhibitor binding loops show remarkable structural conservation from P_2–P'_2 but begin to diverge outwards from this point. In B, a LIGPLOT [78] schematic of the interactions made by the inhibitor PSTI to trypsinogen (databank code 1TGS) are shown which are a common feature of this recognition system. The catalytic serine 195 approaches the carbonyl carbon of the peptide bond under attack while the mainchain amides of residues 193 and 195 forming an oxyanion binding pocket for the peptide carbonyl oxygen. Anti-parallel beta sheet interactions are made from residues (214 and 216 in this system) to the inhibitor backbone on the non-prime side.

the applications of limited proteolysis to the understanding of basic structural data concerning exposure and domain organisation and function will be presented here to illustrate the differing utility and limitations of these approaches. Applications of the technique to ligand-effects and folding/stability will be covered in the following sections.

For proteins of unknown structure, limited proteolysis provides a simple method to gain insights into its tertiary fold. Often the results from such biochemical analyses can be employed to complement modelling studies as was done for the lipase from *Pseudomonas aeruginosa* [32]. In this instance, cleavage at Asp38–Gly39 and Glu46–Val47 by *S. aureus* proteinase V8 confirmed that these residues were surface exposed and this information was incorporated into a 3-dimensional model of the protein. Indeed, the multiple sequence alignment of this lipase to a relative of known structure positioned these residues at surface exposed positions. In other studies, proteolysis by a range of proteinases was employed to gain insights into the structure of the human estrogen receptor ligand binding domain [33]. Extensive limited proteolytic experiments isolated the core binding domain to lie between residues Asn304–Lys529, and suggested that the C-terminal domain from 530–553 is most likely surface exposed.

Proteolysis has also been used as a probe of the structure and dynamics of hirudin from *Hirudinaria manillensis* [34]. The 3-dimensional structure of the relative from *Hirudo medicinalis* was used to help interpret the results from limited proteolysis with V8 proteinase, trypsin, thermolysin and subtilisin, all of which cleave in the region of 41–49. This data demonstrates that like the homologue from *H. medicinalis*, hirudin from *H. manillensis* possesses a well structured N-terminal core domain and a flexible C-terminal loop that is readily cleaved by limited proteolysis.

The understanding of the inhibitory mechanism of serpins, a family of serine proteinase inhibitors that control the proteolytic pathways of blood coagulation, fibrinolysis and inflammation has also been advanced by experimentation using limited proteolysis [35,36]. Serpins ar present in the blood in a number of circulatory forms, of which one latent form involves a partial insertion of the reactive centre loop into a β-sheet. This loop insertion partly protects the loop against cleavage at the hinge region [35]. By forming a binary complex with synthetic peptides, Carrell and co-workers [36] were able to show that the limited proteolytic susceptibility of the reactive site loop could be modified. This provided further evidence for the loop insertion model of the latent forms of the serpins with the loop adopting a helical conformation similar to that found in the non-inhibitory ovalalbumin [37].

Structural information on membrane proteins may also be gained via limited proteolysis experiments, particularly concerning the protein topology with respect to the membrane. This approach relies on the premise that proteases of low specificity will completely degrade the portions of the protein on the cytoplasmic or periplasmic side of the membrane, and is well reviewed by Platt [38]. A recent example of this approach is exemplified by the application of limited proteolysis to the melibiose permease from *E. coli* [39]. Here, not only were transmembrane-connecting loops detected but evidence for their specific roles in substrate binding and catalysis was inferred from differential proteolytic susceptibilities induced by the presence of melibiose and Na^+ and Li^+.

In addition to exploration of surface exposure, limited proteolysis has been widely used to identify and separate individual functional domains within multi-domain proteins [4,5]. Assuming that proteolytic nicking occurs at interdomain linking regions, to determine which domain is responsible for which given function is simply a matter of isolating the products and assaying for activity. This approach depends on the exact domain nature of the protein in question and may not be universally applicable particularly when the domain structure is complex and formed by the amino acid chain crossing between domains several times. Some recent examples are presented here.

The approach is exemplified by studies on the domain structure of a DNA methyltransferase restriction enzyme [40]. Digests performed with trypsin and chymotrypsin have helped identify the regions responsible for specificity, methylation and restriction in this multi-subunit enzyme. The trypsin cleaved enzyme contains two methylation domains and one nicked specificity domain. Similarly, the domain structure of chicken liver xanthine dehydrogenase was predicted by digest with subtilisin [41]. By N-

terminal sequencing of the products, examination of previous sequence analyses and performing activity assays, the authors were able to assign a three domain structure to the protein whereby the 20, 37 and 84 kDa fragments were demonstrat d to contain the iron-sulphur, FAD and molybdenum centres respectively.

A further example of functional assignment is given by proteolysis experiments on pancreatic lipase [42]. The enzyme hydrolyses triglycerides only in the presence of colipase. Limited chymotryptic digestion of the porcine and human protein yielded a stable 12 kDa C-terminal domain that was inactive but was able to bind the co-factor. The N-terminal section was fully degraded. Conversely, digestion of the horse enzyme produces a stable 45 kDa N-terminal fragment that was active but could not properly bind the colipase. Thus limited proteolysis has helped localise the two distinct roles of the enzyme in performing it's function and putatively assign the domains. It is interesting to note here that the same enzyme from different (but closely related) species can possess quite different limited proteolytic susceptibility properties. In the pig and human enzyme, the C-terminal domain remains stable against limited chymotryptic attack whilst in the horse enzyme it is the N-terminal domain. It is quite possible that the ultimate route of digestion is altered by only one or two amino acid changes altering the relative susceptibility of only a few bonds. The resulting proteolytic degradation pathway is altered in favour of one or the other of the two domains being hydrolysed in the different species.

There are many similar examples in the literature, too numerous to mention here, where limited proteolytic fragmentation of a large multi-domain protein has led to an understanding of the different functions of th individual domains. This approach promises to retain its utility now that the identification of high resolution masses has become more routine due to recent advances in mass spectrometry. However, as stated previously, there remain some caveats. Domains are not always formed by contiguous stretches of protein and the concept of "mobile hinges" becomes less applicable. Furthermore, in some instances, loops situated away from domain linking regions may additionally be proteolysed leading to potential misinterpretations. Notwithstanding, limited proteolysis has been widely used to isolate domains not only to identify function, but also to create smaller units for structural analysis by X-ray crystallography or, more increasingly, by NMR.

6. Ligand effects on limited proteolysis

The presence or absence of a bound ligand can profoundly affect the susceptibility of a protein segment to limited proteolysis although the mechanism by which these effects are achieved is not always clear. Indeed, the changes produced may be remote from the site of interaction of the protein with its substrate or co-factor. A number of exemplary cases will be presented here.

The effects that Ca^{2+}-binding can have on proteolytic susceptibility have been known for some time. Indeed, addition of mM concentrations of calcium salts to reaction buffers has long been recognised as a technique to inhibit autolytic degradation of trypsin itself [43]. Presumably calcium is bound to a mobile loop segment, acting as a tether, and reducing the ability of the loop to deform and enter the enzyme's active site. Similarly, calcium ion concentration alters the proteolytic susceptibility of the calcium binding protein calmodulin to endoproteinase Glu-C. The susceptibility of two sites in particular has been monitored and yielded surprising results [44]. A putative cleavage site at Glu87–Ala88 lies close to calcium binding site III and although susceptible in the apo-form, this site is fully protected against cleavage when calcium is bound at this site under the same reaction conditions. More surprising, is the change in susceptibility of Glu31–Leu32 which is situated in calcium binding site I and is resistant to proteolysis in the apo-protein. Susceptibility was enhanced at this bond when calcium was bound at sites III and IV, with free calcium concentrations in the μM range. However, at concentrations above 100 mM and in otherwise the same conditions, full protection against attack was returned when calcium was presumably bound at site I and possibly also site II. These proteolytic "footprinting" experiments demonstrate that calmodulin must occupy at least three conformational substrates dependent on its calcium-bound state.

Calcium-binding loops serve as a typical model for ligand-induced effects on proteolytic susceptibility.

Recent work grafted a calcium-binding loop onto an existing protein via protein engineering and tested the subsequent proteolytic properties of the mutant protein [45]. A surface loop in the neutral protease from *Bacillus subtilis* was replaced by a longer loop from the homologous enzyme thermolysin which is already known to bind calcium. The mutant enzyme was demonstrated to be able to bind calcium, albeit weakly. This was sufficient, however, to diminish autolysis of the neutral protease which also demonstrated increased kinetic thermal stability in solutions containing 0.1 M $CaCl_2$. This strongly suggests that the engineered loop is protected from autolytic degradation by calcium binding.

Another recently studied ligand-binding system is the cellular retinoid-binding proteins which are inherently more susceptible to tryptic cleavage at a single site in the apo form [46]. A model was proposed whereby the ligand binding site is "capped" by the movement of a helical segment after ligand binding. The susceptible site (Arg30–Lys31) is contained in this region. The extra interactions between the largely hydrophobic ligand and the protein anchor the loop to the protein bulk and restrict local segmental mobility.

Hydrophobic interactions are not the only forces that tether loops. Recent studies on the avidin-biotin system highlighted electrostatic interactions between a loop and a bound ligand which accomplish a similar result [47]. The loop between strands 3 and 4 of the avidin calyx (3–4 loop) makes a pair of hydrogen bonds with one of the valerate oxygens of the biotin ligand in the holo-form of the protein. The ligand-bound form is entirely refractive to proteolysis under the same conditions. These interactions are obviously lost in the apo-form which is cleaved by proteinase K specifically at two sites, Thr40–Ser41 and Asn42–Glu43, within the 3–4 loop. Although this result is easily rationalised, a more surprising effect on proteolytic susceptibility was observed with an alternative ligand. The chromatogenic reporter 4′-hydroazobenzene-2-benzoic acid (HABA) binds to the same site as biotin and in the same mode; the X-ray crystal structure of this complex is known [48]. However, HABA lacks a corresponding polar group to valerate in biotin and the hydrogen bonds with the loop are lost. Most surprisingly, the rate of proteolysis of this loop by proteinase K is enhanced, almost by an order of magnitude, with respect to hydrolysis of the wild-type apo-enzyme. This result raises some important issues concerning loop dynamics and proteolysis. How does a *lack* of tethering *increase* proteolytic susceptibility? Perhaps the HABA ligand induces the loop to protrude more from the protein surface increasing its likelihood of being accessed by the attacking enzyme? Similarly, the reduced interactions between the loop and the protein might make the loop more flexible allowing it to take up more conformations. This would increase the frequency with which the loop passes through a conformation acceptable to the active site of the proteinase. Of course, all this speculation ignores the enzyme itself. It is quite possible that the enzyme may also exhibit some inductive effects on loop dynamics. These questions remain unanswered, not least because the 3–4 loop in the crystal structure of the HABA holo-form is disordered and invisible to the crystallographers [48]. Nevertheless, this system shows how proteolytic susceptibility can be modified by different ligands in quite different ways.

7. Limited proteolysis as a probe of unfolding/refolding

Although proteolytic susceptibility is not the sole determinant of protein thermal stability the two are undoubtedly linked [49]. The general observation, based on the early experiments of Linderström–Lang [18], underpins much of the experimentation carried out on protein (un)folding using proteases as structural probes. When a protein is destabilised, either by heating or by chemical denaturation, the resulting unfolding must increase the inherent susceptibility of the protein to proteolytic attack. If the unfolding is global, then the proteolysis will be general and the protein is degraded to completion. This is an "all-or-nothing" proteolytic event; either the protein is degraded to completion or not at all. However, if one segment of the protein is unfolded locally without substantially affecting the overall fold, it may be cut in a limited manner and the nicked protein is stable and remains resistant to further degradation. In this instance, nicked products will accumulate. If, however, the nicked form is considerably less stable, then global unfolding may occur via this intermediate leading to complete degradation. These processes are

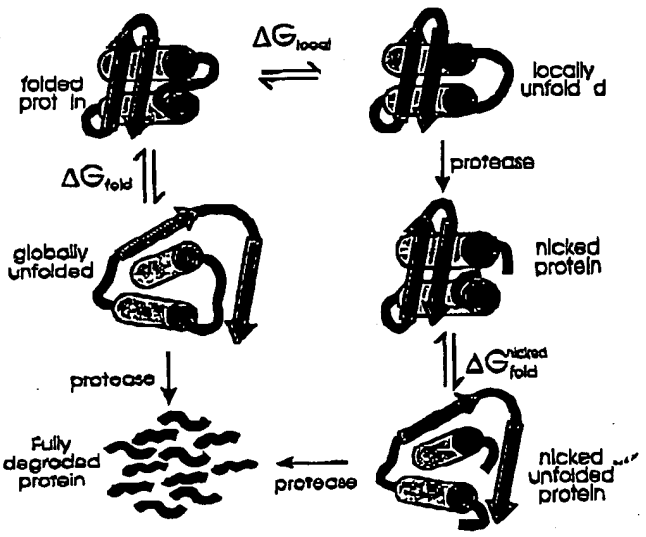


Fig. 4. Unfolding and proteolysis scheme for protein structural studies. For unfolding studies to be successful the protein stability at a given temperature should favour local unfolding above global unfolding ($\Delta G_{local} < \Delta G_{fold}$) otherwise complete protein degradation occurs when the temperature increases. Similarly, if the nicked species is not destabilised significantly ($\Delta G^{nicked}_{fold} \approx \Delta G_{fold}$) then the proteolysis products may be observed. If the nicked species is significantly destabilised, globally unfolding occurs and the protein is degraded to completion. The native folded protein should ideally be refractive to proteolysis under "native" conditions, so that partly folded intermediates may be induced via heating or chemical denaturants.

summarised in Fig. 4. For this type of experiment to be successful, the protease should not readily cleave the native protein (or sub-domain) and the rate of proteolysis should be much faster than the rate of (local) unfolding. Similarly, the amount of energy required for global unfolding should be much larger than for local. If these conditions are satisfied, at elevated temperatures, local unfolding of some segments may be induced and can then be studied using proteases as probes of these partly-(un)folded states. Given that protein unfolding is simply the reverse process of protein folding, inferences may be made on the latter stages of folding.

The autolytic degradation of thermolysin-like proteases represents a classic example of the thermal unfolding model, where in this case the protein is also the attacking protease [50,51]. At elevated temperatures these proteins become irreversibly inactivated due to proteolysis caused by partial (local) unfolding processes [50]. The thermal stability (and hence proteolytic stability) of the protein is largely due to a small surface region and the enzyme can be stabilised by a few mutations in this putative autolytic site [52].

The unfolding of ribonuclease has also been probed by proteinases, using both heating and chemical denaturants to induce partly folded states [53–56]. Early studies monitored the degradation products of the protein across a temperature range from 20°C to 65°C digesting with chymotrypsin [53] and trypsin [54]. The region between residues 17 and 25 is labile above 35°C since chymotrypsin selectively cleaves at Tyr25–Cys26 although the native protein is resistant to chymotrypsin at room temperature. Similarly, trypsin was able to cleave at Lys31–Ser32 and Arg33–Asn34 after heating. Some regions of the protein were more stable: no cleavage was observed between Asn62–Gln74, Ile81–Thr87 and Ala96–Ala102 even at 60°C. More recently, the trypsin results were confirmed by Arnold and co-workers [55] who also found that thermolysin could cleave ribonuclease at Asn34–Leu35 and Thr45–Phe46 at and above 50°C. The results of all these studies suggest that ribonuclease folding elements have varying levels of stability. The mobile loop around residue 20 is readily cleaved by most proteinases even at 20°C. At slightly higher temperatures this region is extended out towards residue 25 where chymotrypsin may cleave. At higher temperatures still, the end of the helix from Asn24–Asn34 must be partly unfolded to allow cleavage by trypsin at this region, and must also destabilise the following β-strand to permit hydrolysis by thermolysin at residue 45. Ribonuclease unfolding has also recently been studied by denaturation with guanidinium chloride [56]. After 5 h of digestion in 1 M guanidinium chloride trypsin cleaves at Arg33–Asn34, similarly to thermal denaturation experiments, and additionally at Arg10–Gln11.

The refolding of ribonuclease has also been studied using urea as a denaturing agent by monitoring tryptic digestion products over time as the urea concentration is lowered by dilution [57]. The putative sites at Lys31–Ser32 and Arg33–Asn34 remained susceptible until the latter stages of refolding. This is in agreement with the unfolding data previously discussed which suggests this region is the last to fold.

A number of recent studies have focused on the limited proteolysis of proteins in the presence of trifluoroethanol (TFE) which is known to favour the formation of α-helix [58–60]. In solution, tertiary structure is believed to be lost in favour of increased α-helical content as demonstrated by CD and protein NMR experiments [61] and hence the protein adopts a partially folded state. Thermolysin has usually been used due to its stability in TFE. When ribonuclease A is incubated in 50% TFE, thermolysin primarily cuts at Asn34–Leu35 although the native protein is resistant to proteolysis [58]. Similarly, normally resistant lysozyme is selectively nicked at Lys97–Ile98 by thermolysin in 50% TFE [59]. Conversely, horse heart cytochrome *c* is fully degraded by thermolysin to many small peptides in aqueous buffer whilst only limited nicking occurs in 50% TFE at the Gly56–Ile57 bond [60]. These experiments have demonstrated the utility of combining proteolysis and TFE to study the partially folded states and complement other biophysical techniques in the study of protein folding.

α-lactalbumin can be unfolded by acid denaturant and by removal of the single bound calcium with a chelating agent [62,63]. In both instances, a partially folded state similar in character to the "molten globule" was induced [64]. This form demonstrated enhanced susceptibility to pepsin, chymotrypsin and proteinase K at essentially the same region (Ala40 to Phe53) although the two domains either side of this segment were resistant. This suggests that despite its molten globule classification, these partly folded states appear to retain significant native-like structure in partially denaturing conditions.

Further recent studies on apomyoglobin have shown that the F-helix becomes distorted in the apoprotein as it is susceptible to a range of proteinases whilst the holo-protein (heme-bound) remains resistant to proteolysis under the same conditions [65]. Additional minor nicking of the apo-form is observed at the B-helix which is also believed to be somewhat mobile. The data strongly suggests that despite the limited proteolysis, the overall native state remains intact and the apo-protein cannot be described as true molten globule. Rather, partial unfolding of the F-helix (and to some extent the β-helix) is responsible for the increased dynamic properties of selected regions and subsequent pattern of proteolysis. This data is in agreement with previous computational and spectroscopic data on the folding of apo-myoglobin which suggest that the F-helix is the last to fold [66–68].

8. Prediction of limited proteolytic sites

Limited proteolysis may also be expressed as a prediction problem. Given a protein structure or sequence and an attacking enzyme, where will the first sites of hydrolysis be located? The ability to achieve either of these is clearly of great benefit. Knowledge of the most susceptible site will permit rational redesign via protein engineering to enhance or reduce proteolytic susceptibility, particularly if the structure is known. Similarly, successful prediction from sequence will inform on the likely structure, assist modelling and provide testable hypotheses. However, the prediction of limited proteolysis has been addressed by only a few groups, and has usually focused on a single system of interest. In a group of 2 S albumins from *Brassica napus* a strong sequence propensity to form a β-turn was noted at proteolytic processing sites [69]. This suggests an elementary prediction scheme for these proteins. A similar scheme was developed to predict the likely processing sites in prohormonal proteins via a probabilistic scheme to predict potential Ω-loops [70]. These are regions containing no regular secondary structure, between 6 and 16 residues in length, and with an end-to-end distance less than 10 Å.

The above studies apply only to very limited systems. Clearly a global approach to the problem is required and a useful first step would be the ability to predict the sites of initial proteolysis in proteins of known tertiary structure. To achieve this goal the conformational analysis of a set of tryptic nicksites in proteins of known structure [26] was expanded to include additional conformational parameters and sites cut by proteinases of broader specificity in order to derive a prediction algorithm. A description of these parameters and the reasons for their consideration is listed in Table 1. Preliminary analyses and modelling has already shown that these parameters are correlated with limited proteolysis [26,28,71]. A thorough analysis revealed that no single conformational parameter is a perfect indicator of limited proteolysis and that a uniformly successful prediction could only

Table 1
Conformational parameters and their relationship to limited proteolysis

Parameter	Reason for inclusion	Method/technical details
Solvent accessibility	Nicksites already known to be generally at surface exposed sites	A probe rolled around the protein exterior assigns areas in $Å^2$ to each atom which are summed over each residue [73].
Protrusion index	Nicksites would be expected to protrude from the protein surface to enable accommodation into the enzyme active site	Each amino acid is assigned a score from 0 to 9 depending on it's protrusion index calculated from a set of similar equimomental ellipsoids with origins at the protein centre of mass [74].
Temperature factors	Nicksites are found at flexible regions of the protein, as characterised by temperature factors or B-values from X-ray crystallographic determinations	The mean residue temperature factor is calculated by summing and averaging individual atomic values.
Ooi numbers	Nicksites would be expected to be at weakly packed regions of the structure which are able to locally unfold more easily	Each amino acid is assigned an Ooi number score which is simply the number of α-carbon centres within a fixed radius [75]. Scores are normalised and subtracted from unity so that higher scores are more favourable.
Secondary structure	Nicksites are rarely found in regular secondary structure and there are geometrical and energetic reasons for their exclusion from helix and sheet	Each amino acid is assigned to one of the three secondary structure states helix, strand or coil [76]. Each state is assigned a score, with an additional penalty
Hydrogen bonding	Nicksites would be expected to be at regions of the structure which are not overly pinned down by interactions with the bulk of the protein such as hydrogen bonds	The number of non-local hydrogen bonds is calculated from residues within a fixed window about each amino acid to those outside it. Values are normalised and subtracted from unity so that higher scores are more favourable.

be achieved when they are all combined in a weighted prediction scheme [72]. Additionally, since limited proteolytic sites typically require 10 residues or more to unfold locally, smoothing windows were applied to the parameter values. A Monte Carlo optimisation procedure was implemented to produce the most favourable combination of weights and smoothing windows with which to combine the parameters. This was based on their ability to discriminate the true nicksite bonds from putative (but not cleaved) sites and all peptide bonds. A set of weights and smoothing windows were found that correctly predicted the first site of nicking for every protein in a data set cut by narrow specificity proteinases (trypsin, V8-proteinase, and endoproteinases Arg-c and Lys-c). Some representative prediction profiles are illustrated in Fig. 5. The true sites of first proteolysis are invariably at peaks in the prediction profile and one of the nicksites is top scoring for each protein. The results were not as good for proteins where data was only available for broad specificity enzymes such as subtilisin and thermolysin. In these cases however, the nicksites were always near the top of peaks and in all but a few cases the top scoring residue was within 2 or 3 amino acids of the first cut site.

An overall conclusion from the analysis is that, although all are important, the types of conformational determinants of limited proteolysis can be broken down into 3 main categories:

1. Flexibility. The ability to unfold locally is critical for limited proteolysis to take place and regions with high segmental mobility are better placed to accomplish this. This is characterised by X-ray temperature factors.

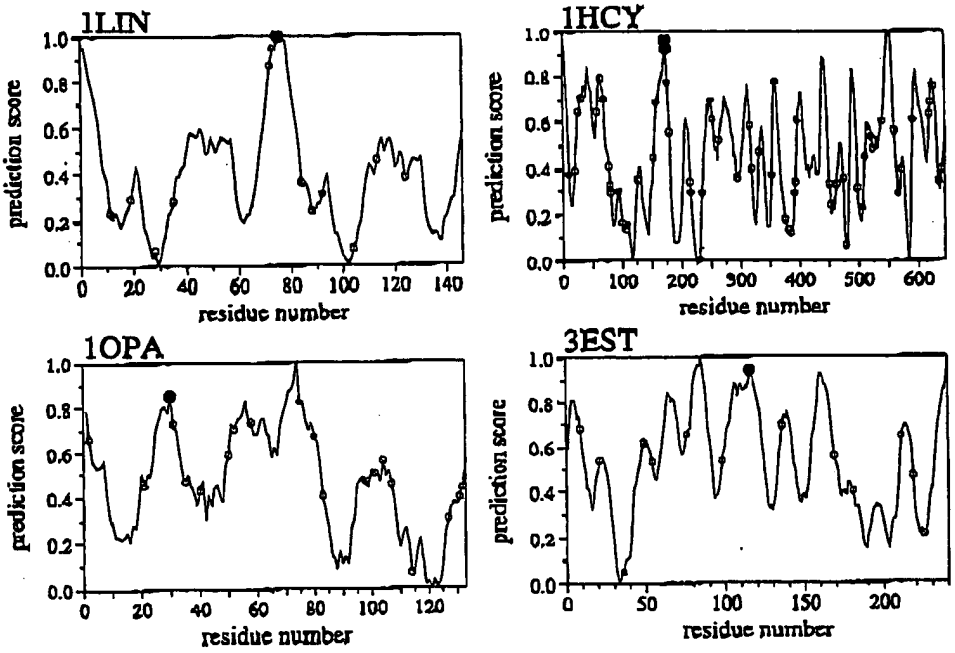


Fig. 5. Limited proteolysis prediction profiles. Prediction profiles are illustrated for four example proteins, calmodulin (Brookhaven databank code [79], 1LIN), hemocyanin (1HCY), cellular retinol-binding protein II (1OPA) and elastase (3EST). The true nicksites are indicated as filled circles on the profiles whilst putative sites matching the primary sequence specificity are indicated by empty circles. The profiles were calculated using the Nickpred program [72] using the optimised weights and smoothing windows for narrow-specificity proteinases.

2. Exposure. Although not an absolute necessity, sites near the surface will be able to undergo local unfolding more easily. Similarly, the conformational change required at surface sites is likely to be less since they are already partly protruding from the surface and hence are more readily accommodated in to an enzyme's active site without causing intermolecular steric hindrance. This is characterised by accessibility, protrusion and Ooi numbers.
3. Local interactions. A good candidate for local unfolding and adaptation must not be unduly restrained by interactions such as hydrogen bonds, disulphide linkages and van der Waals interactions. This is characterised by hydrogen bonding, secondary structure and Ooi numbers.

It is the requirement to unfold locally that appears to be the key determinant, which will be correlated to all of these parameters to a greater or lesser extent. Certainly, predictions are improved if the conformational parameters are considered together in a weighted scheme rather than apart.

9. Future directions

Given that prediction of limited proteolysis from structure is generally possible, the challenge remains to generate a method able to accomplish this from sequence alone. However, the structural determinants of the molecular recognition process are still not completely determined. The precise degree of local unfolding required to facilitate limited proteolysis has yet to be determined. Similarly, the precise degree to which unfolding polypeptides must mimic the idealised substrate binding of proteinase protein inhibitors needs to be quantified more precisely. With this information it may be possible to rationally design mutations that enhance or reduce susceptibility to proteolytic attack and, by inference, the fundamental stability of proteins.

As yet a complete understanding of the dynamics of limited proteolysis is missing. Given the varied conditions under which limited proteolysis is conducted, little comparative data is available. Similarly, the majority of experimentalists view the limited

proteolytic process as a means to an end and do not calculate rate constants. However, the rate of limited proteolysis is an important factor as the process is undoubtedly not a binary "cut or not cut" problem. Indeed, in systems where the unfolding of the protein is much slower than the proteolysis the possibility of directly measuring rates for local unfolding exists. The understanding of protein folding has clearly benefited from the use of proteinases as structural probes and their continued use promises to reveal even more about the partially folded states of proteins.

With a more complete understanding of the molecular recognition events surrounding limited proteolysis and an ability to predict changes upon alteration of the substrate, rational redesign of substrates may be possible. This has obvious consequences for biotechnological uses: increasing the proteolytic stability of a protein such as for therapeutic purposes, or conversely, reducing to give a protein a more limited life span. However, even with current understanding limited proteolysis is still a powerful, and perhaps undervalued, tool for extracting structural information about protein systems.

Acknowledgements

The author would like to acknowledge support from the Wellcome Trust in the form of an Advanced Retraining Fellowship and would like to thank Professor Rob Beynon for invaluable comments and support in the writing of this review.

References

[1] H. Neurath, K.A. Walsh, Proc. Natl. Acad. Sci. U.S.A. 73 (1976) 3825–3832.
[2] M. Ottensen, Ann. Rev. Biochem. 36 (1967) 55–76.
[3] J.S. Bond, R.J. Beynon, Mol. Asp. Med. 9 (1987) 175–287.
[4] H. Neurath, in: R. Jaenicke (Ed.), Protein Folding, 1980, pp. 501–504.
[5] H. Neurath, Chem. Scripta. 27B (1986) 221–229.
[6] F.M. Richards, P.J. Vithayathil, J. Biol. Chem. 234 (1959) 1459–1464.
[7] A. Fontana, P. Polverino de Laureto, V. De Fillipis, E. Scaramella, M. Zambonin, Fold. Des. 2 (1997) 17–26.
[8] D.M. Blow, Acc. Chem. Res. 9 (1976) 145–152.
[9] A. Ruhlmann, D. Kukla, P. Schwager, K. Bartels, R. Huber, J. Mol. Biol. 77 (1973) 417–436.
[10] J. Kraut, Ann. Rev. Biochem. 46 (1977) 331–358.
[11] I. Schechter, A. Berger, Biochem. Biophys. Res. Commun. 27 (1967) 157–162.
[12] W. Lu, I. Apostol, M.A. Quasim, N. Warne, R. Wynn, W.L. Zhang, S. Anderson, Y.W. Chang, E. Ogin, I. Rothberg, K. Ryan, M. Laskowski Jr., J. Mol. Biol. 266 (1997) 441–461.
[13] W. Bode, E. Meyer, J.C. Powers, Biochemistry 28 (1989) 1951–1963.
[14] H. Grøn, M. Meldal, K. Breddam, Biochemistry 31 (1992) 6011–6018.
[15] M. Rheinnecker, G. Baker, J. Eder, A.R. Fersht, Biochemistry 32 (1993) 1199–1203.
[16] Y. Hu, C. Fenwick, A.M. English, Inorg. Chim. Acta 242 (1996) 261–269.
[17] K.L. Bennett, S.V. Smith, R.J.W. Truscott, M.M. Sheil, Anal. Biochem. 245 (1997) 17–27.
[18] K. Linderström-Lang, Cold Spring Harbor Symp. Quant. Biol. 14 (1949) 117–126.
[19] A. Fontana, G. Fassina, C. Vita, D. Dalzoppo, M. Zamai, M. Zambonin, Biochemistry 25 (1986) 1847–1851.
[20] C. Vita, D. Dalzoppo, A. Fontana, Biochemistry 24 (1985) 1798–1806.
[21] H. Taniuchi, C.B. Anfinsen, J. Biol. Chem. 243 (1968) 4778–4786.
[22] H. Taniuchi, C.B. Anfinsen, A. Sodja, Proc. Natl. Acad. Sci. U.S.A. 58 (1967) 1235–1242.
[23] J. Novotný, R.E. Bruccoleri, FEBS Lett. 211 (1987) 185–189.
[24] A. Fontana, in: A. Kotyk, J. Skoda, V. Paces, V. Kostka (Eds.), Highlights of Modern Biochemistry, VSP International Science Publishers, Zeist, The Netherlands, 1989, pp. 1711–1726.
[25] C. Vita, D. Dalzoppo, A. Fontana, in: I.M. Chaiken, E. Chiancone, A. Fontana, P. Neri (Eds.), Macromolecular recognition: Principles and biotechnological applications, Humana Press, Clifton, New Jersey, 1988.
[26] S.J. Hubbard, S.F. Campbell, J.M. Thornton, J. Mol. Biol. 220 (1991) 507–530.
[27] W. Bode, R. Huber, Eur. J. Biochem. 204 (1992) 433–451.
[28] S.J. Hubbard, F. Eisenmenger, J.M. Thornton, Protein Sci. 3 (1994) 757–768.
[29] C. Ghelis, M. Tempete-Gaillourdet, J.M. Yon, Biochem. Biophys. Res. Commun. 84 (1978) 31–36.
[30] J.M. Thornton, M.S. Edwards, W.R. Taylor, D.J. Barlow, EMBO J. 5 (1986) 409–413.
[31] N.C. Price, C.M. Johnson, in: R.J. Beynon, J.S. Bond (Eds.), Proteolytic enzymes-A practical approach, IRL Press, Oxford, UK, 1989, pp. 163–180.
[32] K.-E. Jaeger, S. Ransac, H.B. Koch, F. Ferrato, B.W. Dijkstra, FEBS Lett. 332 (1993) 143–149.
[33] D.A. Seielstad, K.E. Carlson, P.J. Kushner, G.L. Greene, J.A. Katzenellenbogen, Biochemistry 34 (1995) 12605–12615.
[34] A. Vidigni, V. De Filippis, G. Zanotti, C. Visco, G. Orsini, A. Fontana, Eur. J. Biochem. 226 (1994) 323–333.
[35] A.E. Mast, J.J. Enghild, G. Salvesen, Biochemistry 31 (1992) 2720–2728.

[36] W.-S.W. Chang, M.R. Wardell, D.A. Lomas, R.W. Carrell, Biochem. J. 314 (1996) 647–653.
[37] P.E. Stein, A.G.W. Leslie, J.T. Finch, W.G. Turnell, P.J. McLaughlin, R.W. Carrell, Nature 347 (1990) 99–102.
[38] J.M. Pratt, in: R.J. Beynon, J.S. Bond (Eds.), Proteolytic enzymes - A practical approach, IRL Press, Oxford, UK, 1989, pp. 181–191.
[39] C. Gwidzek, G. Leblanc, M. Bassilana, Biochemistry 36 (1997) 8522–8529.
[40] M. Webb, I.A. Taylor, K. Firman, G.G. Kneale, J. Mol. Biol. 250 (1995) 181–190.
[41] A. Sato, T. Nishino, K. Noda, Y. Amaya, T. Nishino, J. Biol. Chem. 270 (1995) 2818–2826.
[42] A. Abousalham, C. Chaillan, B. Kerfelec, E. Foglizzo, Protein Eng. 5 (1992) 105–111.
[43] T. Vajda, A. Garai, J. Inorg. Biochem. 15 (1981) 307–315.
[44] S. Pedigo, M.A. Shea, Biochemistry 34 (1995) 1179–1196.
[45] S. Toma, S. Campagnoli, I. Margarit, R. Gianna, G. Grandi, M. Bolognesi, V. De Filippis, A. Fontana, Biochemistry 30 (1991) 97–106.
[46] R.S. Jamison, M.E. Newcomer, D.E. Ong, Biochemistry 33 (1994) 2873–2879.
[47] D. Ellison, J. Hinton, S.J. Hubbard, R.J. Beynon, Protein Sci. 4 (1995) 1337–1345.
[48] O. Livnah, E.A. Bayer, M. Wilchek, J.L. Sussman, FEBS Lett. 328 (1993) 165–168.
[49] D.A. Parcell, R.T. Sauer, J. Biol. Chem. 264 (1989) 7590–7595.
[50] B. Van den Burg, V.G.H. Eijsink, B.K. Stulp, G.' Venema, Biochem. J. 272 (1990) 93–97.
[51] V.G.H. Eijsink, R. Veltmann, W. Aukema, G. Vriend, G. Venema, Nat. Struct. Biol. 2 (1995) 374–379.
[52] B. Van den Burg, B.W. Dijksira, G. Vriend, B. Van der Vinne, G. Venema, V.G.H. Eijsink, Eur. J. Biochem. 220 (1994) 981–985.
[53] J.A. Rupley, H.A. Scheraga, Biochemistry 2 (1963) 421–431.
[54] T. Ooi, J.A. Rupley, H.A. Scheraga, Biochemistry 2 (1964) 432–437.
[55] U. Arnold, K.P. Rücknagel, A. Schierhorn, R. Ulbrich-Hofman, Eur. J. Biochem. 237 (1996) 862–869.
[56] H.J. Yang, C.L. Tsuo, Biochem. J. 305 (1995) 379–384.
[57] K. Lang, F.X. Schmid, Eur. J. Biochem. 159 (1986) 275–281.
[58] P. Polverino De Laureto, E. Scaramella, V. De Filippis, M. Vruix, M. Rico, A. Fontana, Protein Sci. 6 (1997) 1–13.
[59] P. Polverino De Laureto, V. De Filippis, E. Scaramella, M. Zambonin, A. Fontana, Eur. J. Biochem. 230 (1995) 779–787.
[60] A. Fontana, M. Zambonin, V. De Filippis, M. Bosco, P. Polverino De Laureto, FEBS Lett. 362 (1995) 266–270.
[61] M. Buck, S. Radford, C. Dobson, Biochemistry 32 (1993) 669–678.
[62] P. Polverino De Laureto, V. De Filippis, M. Di Bello, M. Zambonin, A. Fontana, Biochemistry 34 (1995) 12596–12604.
[63] K. Kuwajima, FASEB J. 10 (1996) 102–109.
[64] O.B. Ptitsyn, Curr. Opin. Struct. Biol. 5 (1995) 74–78.
[65] A. Fontana, M. Zambonin, P. Polverino de Laureto, V. De Filippis, A. Clementi, E. Scaramella, J. Mol. Biol. 266 (1997) 223–230.
[66] P.A. Jennings, P.E. Wright, Science 262 (1993) 892–896.
[67] F.M. Hughson, P.E. Wright, R.L. Baldwin, Science 249 (1990) 1544–1548.
[68] J. Tirado-Rives, W.L. Jorgensen, Biochemistry 32 (1993) 4175–4184.
[69] R.I. Monsalve, L. Menéndez-Arias, C. López-Otín, R. Rodriguez, FEBS Lett. 263 (1990) 209–212.
[70] E. Bek, R. Berry, Biochemistry 29 (1990) 178–183.
[71] S.J. Hubbard, J.M. Thornton, S.F. Campbell, Faraday Discuss. 93 (1992) 13–23.
[72] S.J. Hubbard, R.J. Beynon, J.M. Thornton, Protein Eng. (1998) accepted for publication.
[73] B. Lee, F.M. Richards, J. Mol. Biol. 55 (1977) 379–400.
[74] W.R. Taylor, J.M. Thornton, W.G. Turnell, J. Mol. Graph. 1 (1983) 30–38.
[75] K. Nishikawa, T. Ooi, J. Biochem. (Tokyo) 100 (1986) 1043–1047.
[76] W. Kabsch, C. Sander, Biopolymers 22 (1993) 2577–2637.
[77] P. Kraulis, J. Appl. Crystallogr. 24 (1991) 946–950.
[78] A.C. Wallace, R.A. Laskowski, J.M. Thornton, Protein Eng. 8 (1995) 127–134.
[79] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, J. Mol. Biol. 112 (1977) 532–542.